



# Artificial Neural Networks to reconstruct incomplete satellite data: application to the Mediterranean Sea Surface Temperature

E. Pisoni, F. Pastor, M. Volta

## ► To cite this version:

E. Pisoni, F. Pastor, M. Volta. Artificial Neural Networks to reconstruct incomplete satellite data: application to the Mediterranean Sea Surface Temperature. *Nonlinear Processes in Geophysics*, 2008, 15 (1), pp.61-70. hal-00302956

**HAL Id: hal-00302956**

**<https://hal.science/hal-00302956>**

Submitted on 5 Feb 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Artificial Neural Networks to reconstruct incomplete satellite data: application to the Mediterranean Sea Surface Temperature

E. Pisoni<sup>1</sup>, F. Pastor<sup>2</sup>, and M. Volta<sup>1</sup>

<sup>1</sup>Department of Electronic for Automation, University of Brescia, Via Branze 38, 25123 Brescia, Italy

<sup>2</sup>Centro Estudios Ambientales del Mediterraneo, C. Charles Darwin 14, Paterna, Valencia, Spain

Received: 23 July 2007 – Revised: 26 November 2007 – Accepted: 3 January 2008 – Published: 5 February 2008

**Abstract.** Satellite data can be very useful in applications where extensive spatial information is needed, but sometimes missing data due to presence of clouds can affect data quality. In this study a methodology for pre-processing sea surface temperature (SST) data is proposed. The methodology, that processes measures in the visible wavelength, is based on an Artificial Neural Network (ANN) system. The effectiveness of the procedure has been also evaluated comparing results obtained using an interpolation method. After the methodology has been identified, a validation is performed on 3 different episodes representative of SST variability in the Mediterranean sea. The proposed technique can process SST NOAA/AVHRR data to simulate severe storm episodes by means of prognostic meteorological models.

## 1 Introduction

The presence of lacking data is a common problem when working with environmental data. This problem can arise due to insufficient sampling, faults in data acquisition, error in measurements, cloud presence (in the case of satellite data). These issues are even more important when treating data varying spatially and temporally. A lot of approaches have been proposed in literature to process data; by summarizing it is possible to consider the following methodologies:

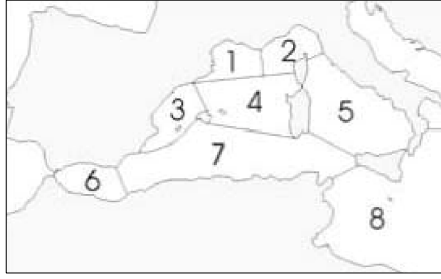
- regression and interpolation model (Acock, 2000, Iglesias et al., 2005, Paatero et al., 2005, Romanowicz et al., 2005);
- neural network and fuzzy logic (Gardner and Dorling, 1998, Nunnari et al., 2004, Kukkonen et al., 2003);
- self organising map (Junninen et al., 2004);

- semi-empirical model (Dirks et al., 2002);
- singular spectrum analysis (Kondrashov and Ghil, 2006);
- empirical orthogonal function (Beckers and Rixen, 2003, Houseago-Stokes and Challenor, 2004);
- expectation maximization algorithm (Schneider, 2001).

In recent years Artificial Neural Networks (ANN) have been widely used, mainly because when correctly trained can approximate roughly any nonlinear function. Some atmospheric science applications show that ANN (Bishop, 1995) give better results than statistical linear methods: for example Perez et al. (2000), Gardner and Dorling (1999), Kolehmainen et al. (2001), Sanz and Marques (2004). Some others (Tangang et al., 1998, Wu et al., 2006) have applied ANN to process tropical Pacific SST.

In this study an ANN based methodology for treating missing data in SST satellite images is proposed, in order to improve the performances of weather forecasting models in extreme meteorological events (such as torrential rains) in the Western Mediterranean Basin. In fact, a previous work (Millan et al., 1995) pointed out that a key factor in the formation of torrential rains in the Spanish East coast is the sea surface temperature (SST) in the Western Mediterranean basin. It has also been shown (Pastor et al., 2001) that the results of mesoscale meteorological models are highly improved when using SST data obtained by satellites in “almost real time”. SST can be efficiently measured with satellite sensor (like NOAA/AVHRR, the Advanced Very High Resolution Radiometer) that builds SST maps from collected radiances using the split-window technique, but a variety of problems can arise from using satellite data. In the case of the SST obtained by NOAA/AVHRR channels 4 and 5, which use atmospheric window observations, a particularly relevant problem is the possible presence of clouds in the sensor field

Correspondence to: E. Pisoni  
(enrico.pisoni@ing.unibs.it)



**Fig. 1.** Domain analyzed in this study with the 8 zones considered.

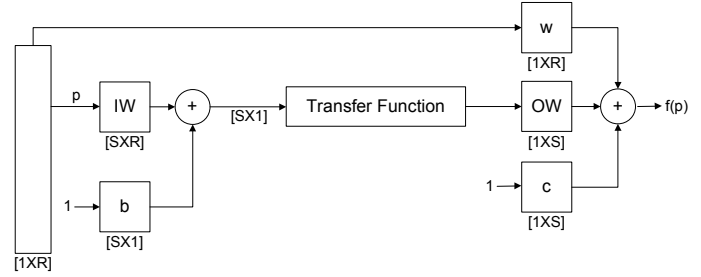
of view. This not only prevents the satellite from measuring radiances; it also impedes the automated (real-time) use of this information in meteorological modeling systems, that cannot accept lacking data. For this reason it's necessary to pre-process SST data.

To perform this study Western Mediterranean basin has been divided from a phenomenological point of view in 8 zones and an ANN system has been designed, identified and validated for each zone and season. An application of the proposed technique will be implemented in the future, pre-processing SST NOAA/AVHRR data of a severe storm episode. This dataset will be then ingested in RAMS meteorological model, and results compared with a RAMS simulation fed with SST standard dataset.

## 2 Methodology and data

In this Section the formulation of the ANN is presented. To apply the proposed technique the selected domain is the Mediterranean Sea, divided into 8 areas (see Fig. 1) characterized by its own and peculiar characteristics, based on geographical, oceanographical and meteorological criteria.

This division is also useful as it is not possible to design a single ANN accounting for the whole SST variability. Another important variable influencing SST variability is seasonality, and so a temporal division in ANN construction was introduced. At the end 32 “categories” (and so 32 ANN structures) are considered, combining 8 spatial and 4 temporal features. In particular the temporal periods considered are DJF (from December to February), MAM (from March to May), JJA (from June to August) and SON (from September to November). To evaluate ANN results, a simple interpolation method is introduced and applied on the same domain. The following part is structured as follows: Sects. 2.1 and 2.2 present ANN and Interpolation methodologies, Sect. 2.3 describes available data and Sect. 2.4 shows indexes used to evaluate the two techniques.



**Fig. 2.** Structure of the cascade – forward ANN used.

### 2.1 Artificial Neural Network

ANN can be used effectively to reconstruct non-linear relationship learning from training data. An ANN model basically consists of processing elements (called neurons) and connections between elements. Every single neuron performs a weighted sum of inputs that receives from neighboring neurons, uses an activation function to manipulate data and passes results to following neurons. Neurons are structured in layers, that can be of 3 different typologies: input layer (where input data enter the network), hidden layer (where intermediate calculations are performed) and output layer (where final results are produced). Different possible ANN structures can be used, depending on the architecture of the network (number of layers, neurons, etc...), the model parameters, the transfer function used. In this study a cascade-forward network with 2 layers is chosen. The structure of this network consists of a) a first layer that has weights coming from the input, and of b) subsequent layers that have weights coming from the input and all previous layers. All layers have biases, while the last layer is the network output. This structure can be formalized as follows:

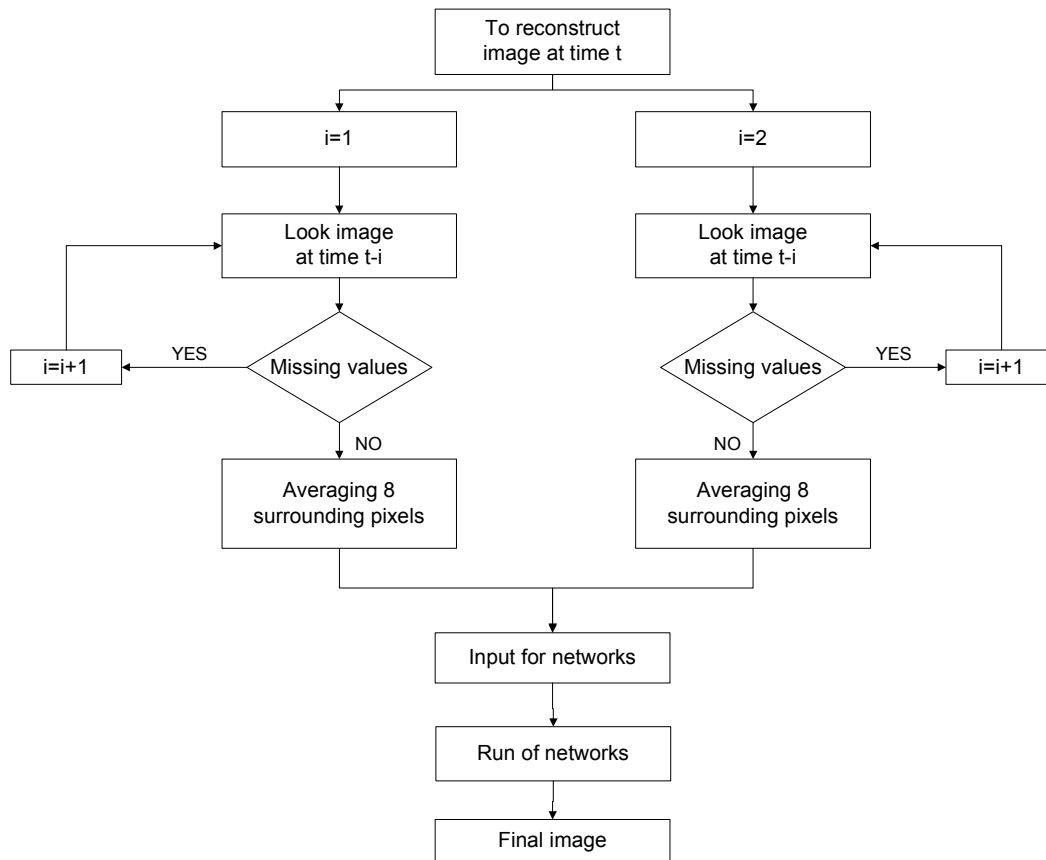
$$f(p) = \sum_{j=1}^S \mathbf{OW}_j \cdot \mathbf{a}_j + \sum_{i=1}^R \mathbf{w}_i \cdot p_i + c \quad (1)$$

with

$$\mathbf{a}_j = TF \left( \sum_{i=1}^R \mathbf{IW}_{j,i} \cdot p_i + b_j \right) \quad (2)$$

where:

- $\mathbf{p}$  represents the input vector;
- $p_i$  represents the input vector element;
- $\mathbf{IW}$  and  $\mathbf{OW}$  are weight matrices of first and second layer;
- $\mathbf{w}$  is the weight vector of second layer coming from first layer input;
- $\mathbf{b}$  and  $c$  are network bias weights;



**Fig. 3.** Algorithm used to reconstruct SST using ANN in the validation phase.

- $S$  is the number of neuron of the first layer;
- $R$  is the size of input vector;
- $TF$  is the used Transfer Function.

Another way to represent the general formalization of ANNs, shown in Eqs. (1) and (2), is depicted in Fig. 2, where the structure of the network is graphically displayed.

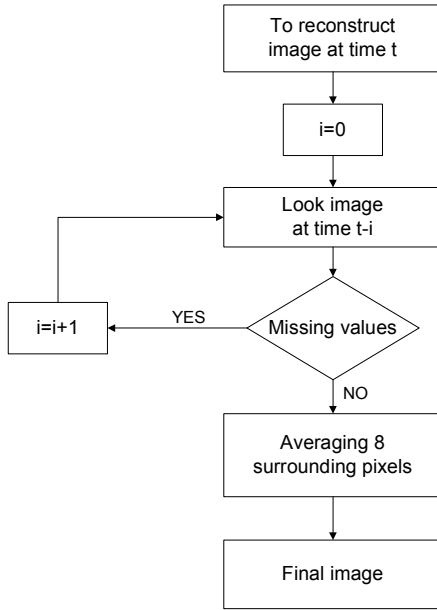
In the considered case study, the target for ANNs is the SST for a particular cell of the domain at time  $t$ . The two input of ANN are the mean values of SST at time  $t-1$  and  $t-2$ ; such values are referred to mean values of the eight pixels surrounding the missing target data. This input-output structure is applied for training phase. For the validation phase, if needed, a preprocess of the input data is performed. In fact, when SST at time  $t-1$  or  $t-2$  are not available due to presence of clouds, their values are reproduced in an iterative way using information at previous time steps. (see Fig. 3). In this way, the ANNs can be more extensively applied over the study domain. This does not happen in the training phase, where the cloud days are not considered.

## 2.2 Interpolation

To evaluate ANN performances an interpolation procedure is introduced (Seze and Desbois, 1987). Due to the fact that SST images are very often characterized by massive presence of clouds, to reconstruct missing data it is necessary to use an interpolation methodology that mix spatial and temporal information. The following algorithm has been implemented:

1. if missing value is surrounded by measures, it is calculated averaging the 8 surrounding pixels;
2. if the previous step can not be applied for all pixels at time  $t$ , an average is performed looking at the surrounding pixels not only at time  $t$ , but also at time  $t-1$ ;
3. the second step is repeated looking more and more in the past till the image is completely reconstructed.

In Fig. 4 a flow-chart of the interpolation algorithm is presented. It's interesting to compare Fig. 3 (ANN method) and Fig. 4 (Interpolation method): the main difference is that in Fig. 4 results of the averaging procedure represents the final value of reconstructed pixels, while in Fig. 3 results of the same procedure are used to create input to the ANN.



**Fig. 4.** Algorithm used to reconstruct SST using interpolation method.

### 2.3 Data

Data used in this study are obtained from the NOAA/NASA AVHRR Oceans Pathfinder dataset, available at PODAAC website ([http://podaac.jpl.nasa.gov/DATA\\_PRODUCT/SST](http://podaac.jpl.nasa.gov/DATA_PRODUCT/SST)) and derived from the 5-channel Advanced Very High Resolution Radiometers (AVHRR) on board the NOAA -7, -9, -11 and -14 polar orbiting satellites.

The available dataset:

- ranges from 1st January 1985 through July 2003, and is derived from interim versions 4 and 4.1 algorithm developed by the University of Miami and explained in Kilpatrick et al. (2001);
- is split in two daily passes (ascending daytime pass and descending nighttime pass), with three different spatial resolutions (54, 18 and 9 km), and two different datasets (“Best pixels” that considers only pixels with high quality flag<sup>1</sup>, and “All pixels” that retains values regardless of their quality flag).

For this study, the descending night-time pass (because not disturbed by sea reflection) and “All pixels” (choosing for the training only data with quality flag bigger than 2) dataset were chosen. The resolution selected is 9 km, that is the best

<sup>1</sup>The SST Pathfinder data have an associated quality flag, ranging from 0 to 3 depending on its success at some quality tests, with the 0 value associated to pixels that do not pass any of the tests and the 3 value associated to pixels that pass all the tests.

**Table 1.** Considered ANN parameter configuration.

Parameter	Configuration
Neuron number in layer 1	4,8
Transfer function of 1st layer	Linear, Log-Sigmoid, Tan-Sigmoid
Neuron number in layer 2	1
Transfer function of layer 2	Linear
Network typology	Cascade forward
Network input	2
Training algorithms	Batch Gradient Descent, Batch gradient Descent with Momentum, Variable Learning, Rate Backpropagation, Resilient Backpropagation, Conjugate Gradient, Quasi-newtonian

**Table 2.** ANN best parameter configuration selected for each period.

Month	Input	Neuron layer 1	Trans. Func.	Training
DJF	2	8	tansig	trainOSS
MAM	2	4	tansig	CGP
JJA	2	8	tansig	GDX
SON	2	4	logsig	GDX

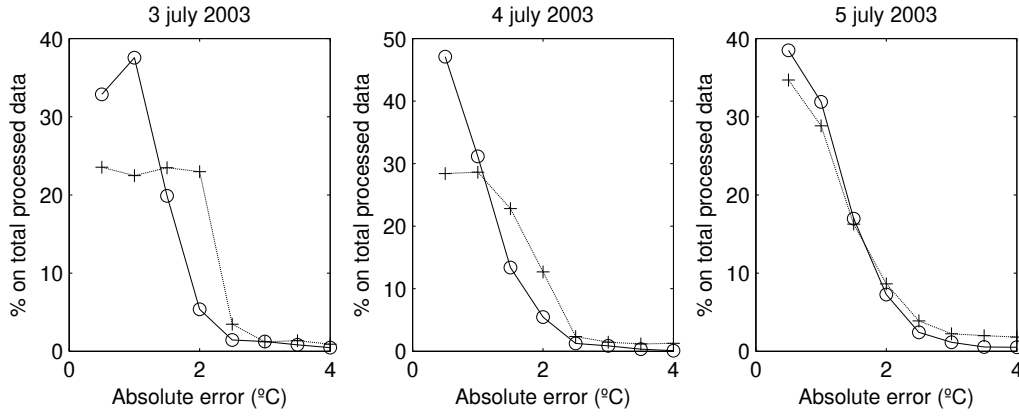
available resolution at the time this study started. The mean percentage of data (for the training period) with quality flag bigger than 2 ranges between 40.2 and 44.8% for the 8 zones. Data with quality flag of 0 or 1 are not used, because often characterised by unrealistic values. The available dataset (as specified spanning from 1985 to 2003) has been divided in training dataset (the period from 1992 to 2001) and validation dataset (selecting in total 3 significant episodes before 1992 and after 2001, as stressed in Sect. 4).

### 2.4 Indexes

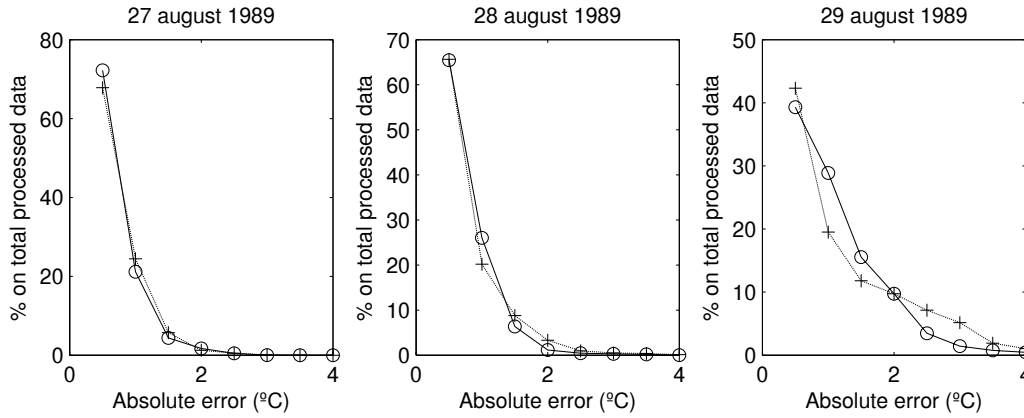
To assess performances of the proposed methodologies, statistical indexes have been used (Legates and McCabe, 1999; Gardner and Dorling, 2000). In particular, indexes used are the Mean Bias Error (MBE), the Root Mean-Square Error (RMSE), the Index of Agreement (IoA) and the Correlation Index (CORR), as defined in the following equations:

$$\text{MBE} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i) \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2} \quad (4)$$



**Fig. 5.** Comparison of absolute error for the 2 analyzed methodologies (ANN with circles, Interpolation with crosses), for the first episode.



**Fig. 6.** Comparison of absolute error for the 2 analyzed methodologies (ANN with circles, Interpolation with crosses), for the second episode.

$$IoA = 1 - \frac{\left[ \sum_{i=1}^n |Y_i - \hat{Y}_i|^2 \right]}{\left[ \sum_{i=1}^n (|\hat{Y}_i - \bar{Y}_i| + |Y_i - \bar{Y}_i|)^2 \right]}$$

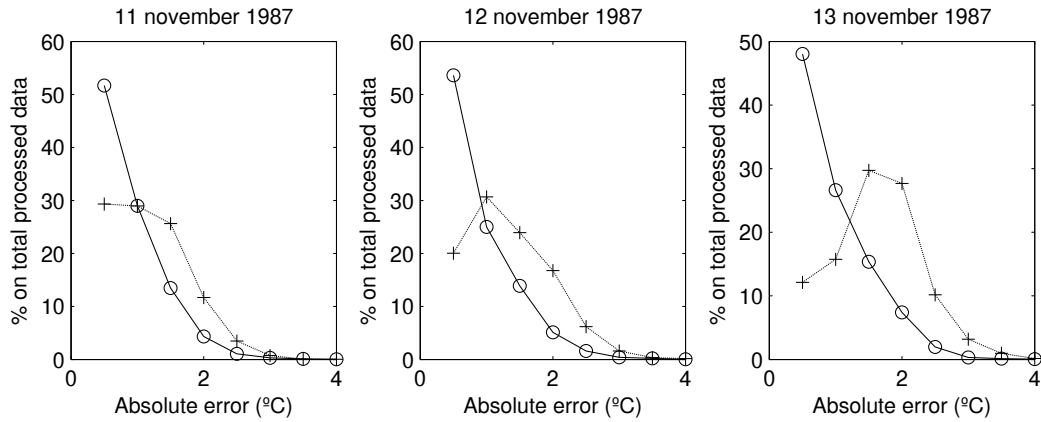
$$CORR = \frac{\sum_{i=1}^n [(Y_i - \bar{Y}_i) \cdot (\hat{Y}_i - \bar{\hat{Y}}_i)]}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2 \cdot \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}}_i)^2}}$$

where

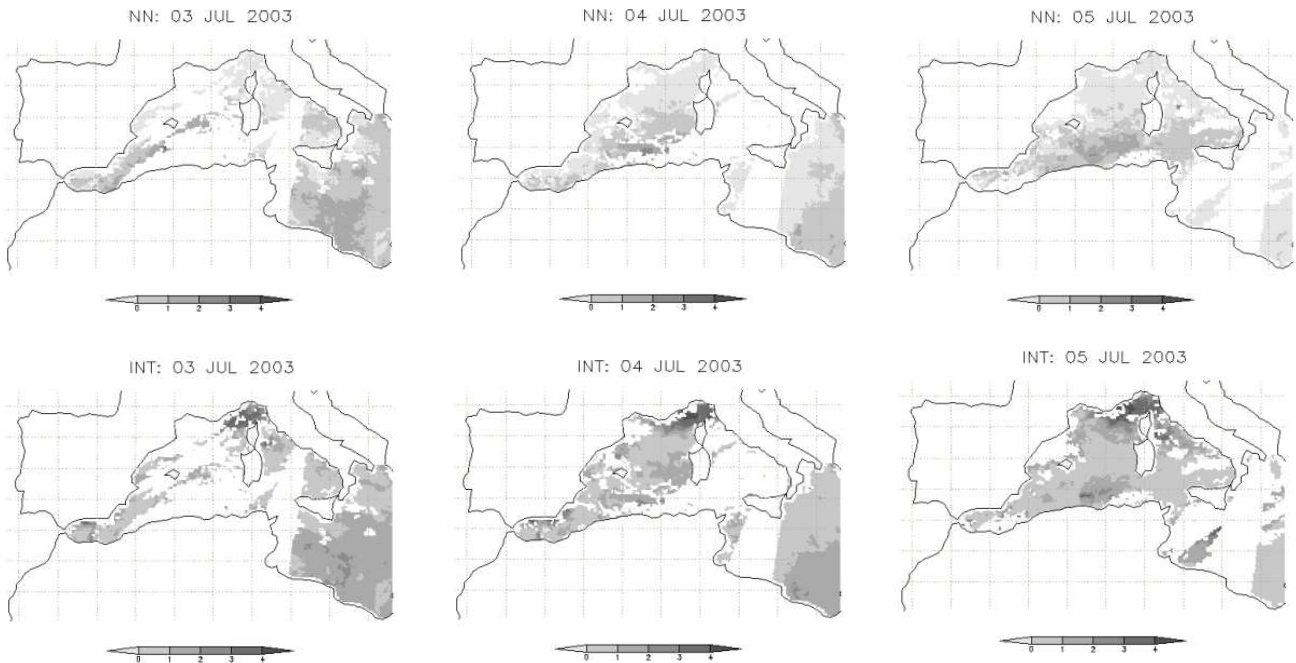
- $Y_i$  is the observed SST;
- $\bar{Y}_i$  is the mean observed SST;
- $\hat{Y}_i$  is the modeled SST;
- $\bar{\hat{Y}}_i$  is the mean modeled SST;

### 3 Training of Artificial Neural Network

- (5) The training of ANN has been performed using parameters configuration shown in Table 1. Considering the training period (the period from 1992 to 2001 has been selected for this task) all the ANNs possible configuration shown in Table 1 have been tested. At the end for each temporal period the best parameter set has been selected, as shown in Table 2.
- (6) Then the best selected structure (for each temporal period) has been extended to the whole geographical domain. In Table 3 correlation values calculated extending the best structures to the whole domain are shown. Table 3 clearly shows good performances of the ANN training phase, with better values in summer time (when more data are available due to cloud absence).



**Fig. 7.** Comparison of absolute error for the 2 analyzed methodologies (ANN with circles, Interpolation with crosses), for the third episode.

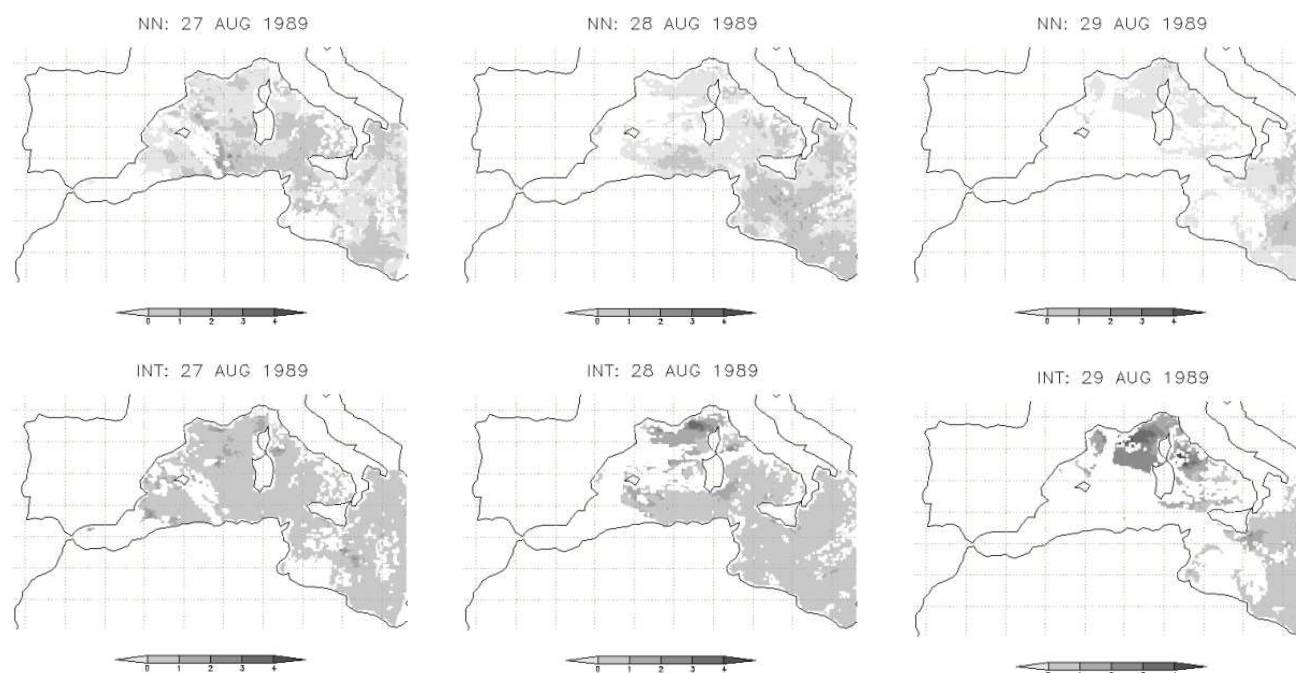


**Fig. 8.** Differences in  $^{\circ}\text{C}$  between ANN method and measured SST (top) and Interpolation method and measured SST (bottom), for the first episode.

#### 4 Validation of Artificial Neural Network and comparison with interpolation method

To validate the ANN methodology, an analysis is performed considering three different episodes, selected to be representative of different Spanish East coast SST patterns. It is necessary to choose episodes not used in the ANN training period; for this reason periods selected to perform validation phase are before 1991 and after 2002, more precisely:

- 3–5 July, 2003 (representative of high SST);
- 27–29 August, 1989 (representative of high SST and close to the 4 September 1989 severe rain episode);
- 11–13 November, 1987 (representative of lower SST and close to the 1 November 1987 severe rain episode).



**Fig. 9.** Differences in  $^{\circ}\text{C}$  between ANN method and measured SST (top) and Interpolation method and measured SST (bottom), for the second episode.

In the following part of the article each period is analyzed performing a comparison between the 2 methodologies previously presented and evaluating statistical indexes. The validation is realized considering only pixels, for the particular selected episode, where measures were available (to be able to compare SST measures and reconstructed data), and for this reason images reconstructed in validation phase (and presented in the following part of the study) are characterized by the presence of “white areas”, that are areas where SST could not be measured due to presence of clouds (see Figs. 8, 9 and 10).

#### 4.1 First episode

The first episode, representative of high SST, was selected in 2003, and is the last temporal period available at PO-DAAC website at the time of this study; in fact after July 2003, a new AVHRR SST algorithm was introduced, and so a comparison with more recent periods is not possible. In Table 4 percentages of no-cloudy pixels (that is to say data used in the validation phase) are shown, for every zone and day of episode. In general, it is possible to notice a good availability of data to perform validation. In Table 5, performances of the ANN methodology are shown. Comparing these performances with the ones obtained with the Interpolation methodology (in the same Table in *italic*), it is possible to appreciate how Mean Bias, RMSE, Index of Agreement and Correlation improve using ANN.

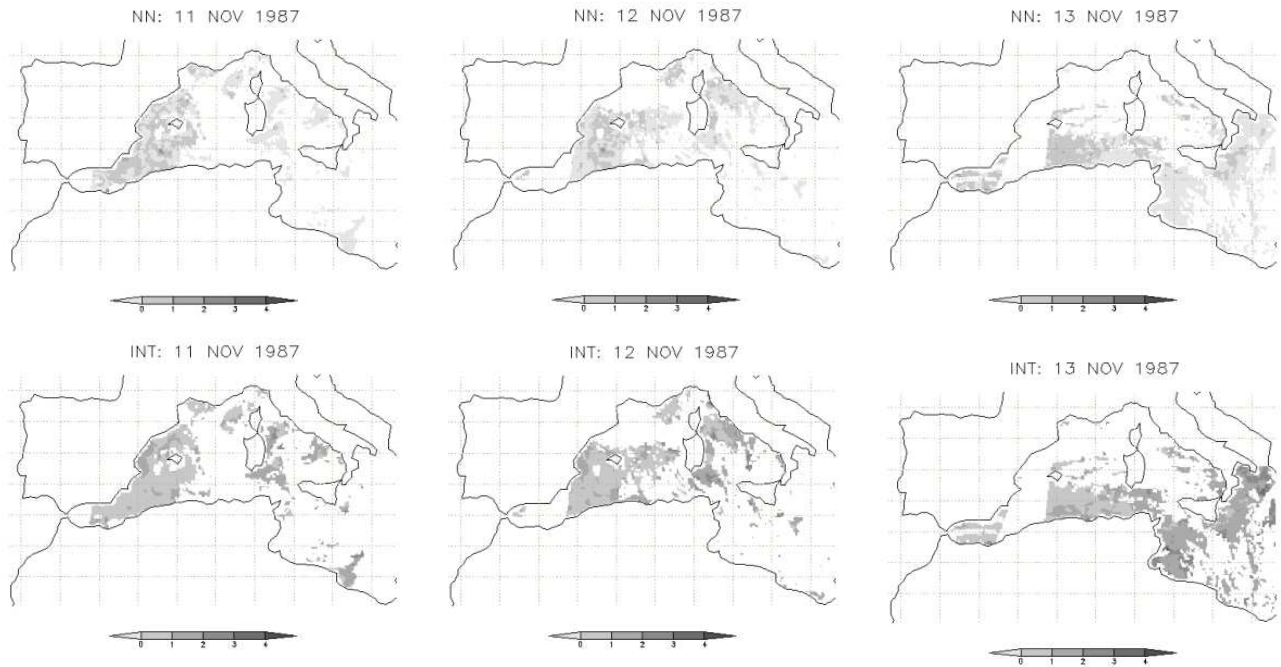
**Table 3.** Correlation values calculated in the training phase for reconstructed areal mean values.

Area	Corr(DJF)	Corr(MAM)	Corr(JJA)	Corr(SON)
1	0.68	0.97	0.94	0.98
2	0.86	0.98	0.95	0.98
3	0.83	0.97	0.94	0.98
4	0.82	0.98	0.96	0.98
5	0.92	0.98	0.96	0.96
6	0.84	0.93	0.89	0.96
7	0.90	0.97	0.97	0.98
8	0.86	0.97	0.98	0.96

Figure 5 shows how absolute error (calculated as absolute difference of simulated and measured temperature) is bigger than 2 degrees only for a very little percentage of data in the ANN case, and that error of ANN methodology drop down faster than Interpolation one, for the majority of the episode (in this Figure, as in Figs. 6 and 7, the y-axis represents the percentage over the total of reconstructed data, for a particular day).

Previous analysis is performed regardless of pixel areal membership. To examine spatial behavior in Fig. 8 maps of differences between measured and reconstructed temperature





**Fig. 10.** Differences in  $^{\circ}\text{C}$  between ANN method and measured SST (top) and Interpolation method and measured SST (bottom), for the third episode.

are depicted, showing how reconstruction of ANN seems to perform better in comparison to Interpolation for the whole domain and period.

#### 4.2 Second episode

Second episode, 27–29 August 1989, is chosen to represent high SST values and to be close to 4 September 1989 severe rain episode. This is a more difficult situation, because in this episode there are a lot of consecutive cloudy days, and so it is a more challenging case for the 2 methods. In Table 4, percentages of non-cloudy pixels are shown; in some zones there is a very small presence of data. Looking at global performances in Table 5 it is possible to notice that performances are better for ANN, except for correlation in the first day of episode, perhaps denoting a difficulty of ANN methodology to correctly describe the involved nonlinear phenomena.

Even if there are small differences in absolute error in Fig. 6 between the 2 methods, when spatially analysing results in Fig. 9, ANN overwhelms the Interpolation method reconstruction.

#### 4.3 Third episode

The third episode is 11–13 November 1987, and is similar to 1989 episode; in fact it's a severe rain episode but with

lower (not too high) SST values. In Table 4 percentages of non-cloudy pixels are shown. It must be considered that in the third day of episode a lot of areas have low percentage of usable data. Here again, as in the first episode, performances of ANN overwhelm Interpolation ones (Table 5).

In Fig. 7, absolute error shows how in the case of ANN, 80% of reconstructed data shows an error that is inferior to one temperature degree; this percentage is lower than the one considering the Interpolation method.

It's worthwhile to note the bad trend of Interpolation in the third day of the episode in Fig. 7, where a strong percentage of data has error between one and two degrees, with a strange gaussian shape. In Fig. 10, it is clearly visible how ANN spatially performs better than Interpolation.

## 5 Conclusions

This investigation is about the implementation of a methodology for pre-processing SST data, to reconstruct missing data due to presence of clouds. The methodology consists of an ANN System that has been identified and validated on a Mediterranean Sea domain. Performances of the methodology have been compared with results of an interpolation method, showing that the ANN system gives better results. At the moment ANN is overestimating SST reconstruction

**Table 4.** Percentage of data used for validation phase, for every zone-day of the three considered episode.

	3 July 2003	4 July 2003	5 July 2003	27 Aug 1989	28 Aug 1989	29 Aug 1989	11 Nov 1987	12 Nov 1987	13 Nov 1987
1	11.5	62.5	85.2	93.9	55.5	54.7	44.6	10.9	6.6
2	79.5	87.2	90.1	79.8	83.6	92.7	45.8	47.5	3.1
3	53.8	76.8	50.1	74.6	13.3	9.9	95.2	77.9	7.1
4	35.8	95.4	96.0	92.2	38.7	36.5	25.7	36.8	23.3
5	64.8	14.1	84.2	85.0	79.7	55.8	39.5	53.1	31.4
6	91.4	92.3	67.1	2.7	0.0	0.8	64.3	12.4	69.4
7	26.6	61.3	94.1	80.6	61.4	3.9	48.1	68.2	65
8	87.2	74.3	36.9	83.8	94.9	63.8	10.5	9.1	61.7

**Table 5.** Global performances of the ANN method (normal font) and Interpolation method (*italic font*) for the three considered episodes.

Index	3 July 2003	4 July 2003	5 July 2003	27 Aug 1989	28 Aug 1989	29 Aug 1989	11 Nov 1987	12 Nov 1987	13 Nov 1987
Mean Bias	0.23	−0.23	−0.35	−0.03	−0.10	−0.69	−0.38	−0.46	−0.49
	<i>0.56</i>	<i>0.08</i>	<i>−0.45</i>	<i>−0.27</i>	<i>−0.25</i>	<i>−0.91</i>	<i>−0.80</i>	<i>−1.04</i>	<i>−1.30</i>
RMSE	1.04	0.93	1.10	0.54	0.61	1.12	0.78	0.86	0.89
	<i>1.51</i>	<i>1.51</i>	<i>1.51</i>	<i>0.57</i>	<i>0.73</i>	<i>1.43</i>	<i>1.07</i>	<i>1.27</i>	<i>1.52</i>
Agreement	0.91	0.94	0.91	0.90	0.88	0.82	0.93	0.88	0.91
	<i>0.79</i>	<i>0.79</i>	<i>0.80</i>	<i>0.89</i>	<i>0.82</i>	<i>0.68</i>	<i>0.87</i>	<i>0.77</i>	<i>0.78</i>
Correlation	0.86	0.90	0.84	0.82	0.85	0.85	0.91	0.85	0.90
	<i>0.70</i>	<i>0.67</i>	<i>0.67</i>	<i>0.84</i>	<i>0.80</i>	<i>0.75</i>	<i>0.90</i>	<i>0.84</i>	<i>0.88</i>

values, but performances could be further improved considering different ANN structures, or i.e. other input to the networks (radiances instead of SST data, etc. ...). A general conclusion from this study is that ANN can be efficiently used to reconstruct SST satellite missing data in the Mediterranean Area, but also that ANN could be a more general tool to process incomplete satellite missing images. This methodology will be soon used to pre-process SST AVHRR/NOAA data, to simulate a severe storm episode with RAMS meteorological model comparing results with a simulation performed with RAMS SST standard dataset. This methodology can also be used in a forecasting modeling system, in which RAMS uses an improved SST dataset, being able, in this way, to improve its weather forecasting.

Edited by: H. A. Dijkstra

Reviewed by: Two anonymous referees

## References

- Acock, M. C.: Estimating Missing Weather Data for Agricultural Simulations Using Group Method of Data Handling, *J. Appl. Meteorol.*, 39, 1176–1184, 2000.
- Beckers, J. M. and Rixen, M.: EOF Calculations and Data Filling from Incomplete Oceanographic Datasets, *J. Atmos. Ocean. Tech.*, 20, 1839–1856, 2003.
- Bishop, C. M.: Neural networks for pattern recognition, Oxford University Press, 1995.
- Dirks, K. N., Johns, M. D., Hay, J. E., and Sturman, A. P.: A simple semi-empirical model for predicting missing carbon monoxide concentrations, *Atmos. Environ.*, 36, 5953–5959, 2002.
- Gardner, M. and Dorling, S.: Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences, *Atmos. Environ.*, 32, 2627–2636, 1998.
- Gardner, M. and Dorling, S.: Neural network modelling and prediction of hourly NO<sub>x</sub> and NO<sub>2</sub> concentrations in urban air in London, *Atmos. Environ.*, 33, 709–719, 1999.
- Gardner, M. and Dorling, S.: Statistical surface ozone models: an improved methodology to account for non-linear behaviour, *Atmos. Environ.*, 34, 21–34, 2000.
- Housego-Stokes, R. and Challenor, P.: Using PPCA to Estimate EOFs in the Presence of Missing Values, *J. Atmos. Ocean. Tech.*, 21, 1471–1480, 2004.
- Iglesias, P., Jorquera, H., and Palma, W.: Data analysis using regression models with missing observations and long-memory: an application study, *Comput. Stat. Data An.*, 2005.
- Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., and Kolehmainen, M.: Methods for imputation of missing values in air quality data sets, *Atmos. Environ.*, 38, 2895–2907, 2004.
- Kilpatrick, K., Podesta, G., and Evans, R.: Overview of the NOAA/NASA Advanced Very High Resolution Radiometer Pathfinder Algorithm for Sea Surface Temperature and Associated Matchup Database., *J. Geophys. Res.*, 106, 9179–9197, 2001.
- Kolehmainen, M., Martikainen, H., and Ruuskanen, J.: Neural networks and periodic components used in air quality forecasting, *Atmos. Environ.*, 35, 815–825, 2001.
- Kondrashov, D. and Ghil, M.: Spatio-temporal filling of missing

- points in geophysical data sets, *Nonlin. Processes Geophys.*, 13, 151–159, 2006,  
<http://www.nonlin-processes-geophys.net/13/151/2006/>.
- Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., and Cawley, G.: Extensive evaluation of neural network models for the prediction of NO<sub>2</sub> and PM<sub>10</sub> concentrations, compared with a deterministic modelling system and measurements in central Helsinki, *Atmos. Environ.*, 37, 4539–4550, 2003.
- Legates, D. and McCabe, G.: Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation, *Water Resour. Res.*, 35, 233–242, 1999.
- Millan, M., Estrela, M., and Caselles, V.: Torrential precipitations on the Spanish east coast: The role of the Mediterranean sea surface temperature, *Atmos. Res.*, 36, 1–16, 1995.
- Nunnari, G., Dorling, S., Schlink, U., Cawley, G., Foxall, R., and Chatterton, T.: Modelling SO<sub>2</sub> concentration at a point with statistical approaches, *Environ. Modell. Softw.*, 19, 887–905, 2004.
- Paatero, P., Aalto, P., Picciotto, S., Bellander, T., Castao, G., Cattani, G., Cyrus, J., Kulmala, M., Lanki, T., and et al., F. N.: Estimating time series of aerosol particle number concentrations in the five HEAPSS cities on the basis of measured air pollution and meteorological variables, *Atmos. Environ.*, 39, 2261–2273, 2005.
- Pastor, F., Estrela, M., Penarrocha, D., and Millan, M.: Torrential Rains on the Spanish Mediterranean Coast: Modeling the Effects of the Sea Surface Temperature, *J. Appl. Meteorol.*, 40, 1180–1195, 2001.
- Perez, P., Trier, A., and Reyes, J.: Prediction of PM<sub>2.5</sub> concentrations several hours in advance using neural networks in Santiago, Chile, *Atmos. Environ.*, 34, 1189–1196, 2000.
- Romanowicz, R., Young, P., Brown, P., and Diggle, P.: A recursive estimation approach to the spatio-temporal analysis and modelling of air quality data, *Environ. Modell. Softw.*, 2005.
- Sanz, B. M. and Marques, N. M.: Total ozone time series analysis: a neural network model approach, *Nonlin. Processes Geophys.*, 11, 683–689, 2004,  
<http://www.nonlin-processes-geophys.net/11/683/2004/>.
- Schneider, T.: Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values, *J. Climate*, 14, 853–871, 2001.
- Seze, G. and Desbois, M.: Cloud Cover Analysis from Satellite Imagery Using Spatial and Temporal Characteristics of Data, *J. Clim. Appl. Meteorol.*, 26, 287–303, 1987.
- Tangang, F., Hsieh, W., and Tang, B.: Forecasting the regional sea surface temperatures of the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors., *J. Geophys. Res.*, 103, 7511–7522, 1998.
- Wu, A., Hsieh, W. W., and Tang, B.: Neural Network forecasts of the tropical Pacific sea surface temperatures, *Neural Networks*, 19, 145–154, 2006.